

AD-A147 835

RULE-BASED EVIDENCE ACCRUAL SYSTEM FOR IMAGE  
UNDERSTANDING(U) HONEYWELL SYSTEMS AND RESEARCH CENTER  
MINNEAPOLIS MN N MARQUINA SEP 84 AFOSR-TR-84-0945

1/1

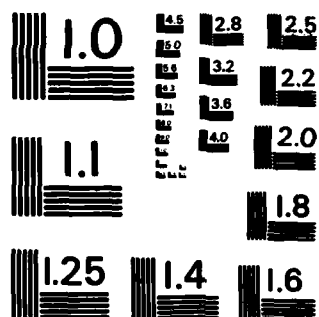
UNCLASSIFIED

F49620-83-C-0134

F/G 20/6

NL





**MICROCOPY RESOLUTION TEST CHART**  
**NATIONAL BUREAU OF STANDARDS-1963-A**

AD-A147 835

## PORT DOCUMENTATION PAGE

②

1a. REPORT SECURITY CLASSIFICATION <b>UNCLASSIFIED</b>		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE		5. MONITORING ORGANIZATION REPORT NUMBER(S) <b>AFOSR-TR- 84-0945</b>	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research	
6a. NAME OF PERFORMING ORGANIZATION Honeywell Systems and Research Center		7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448	
6b. OFFICE SYMBOL (If applicable)		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-83-C-0134	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR		10. SOURCE OF FUNDING NOS	
8b. OFFICE SYMBOL (If applicable) NM		PROGRAM ELEMENT NO. 61102F	
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332-6448		PROJECT NO. 2304	
		TASK NO. A7	
		WORK UNIT NO.	
11. TITLE (Include Security Classification) RULE-BASED EVIDENCE ACCRUAL SYSTEM FOR IMAGE UNDERSTANDING			
12. PERSONAL AUTHOR(S) Nelson Marquina			
13a. TYPE OF REPORT Technical		13b. TIME COVERED FROM _____ TO _____	
		14. DATE OF REPORT (Yr., Mo., Day) SEP 84	
		15. PAGE COUNT 6	
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number.)	
FIELD	GROUP	SUB GR	
19. ABSTRACT (Continue on reverse if necessary and identify by block number.) The main function of an evidence accrual system for image understanding is to sequentially update information on scene objects based on new sensor data or on non-sensory information such as intelligence. This paper presents a concept for sequentially updating information on scene objects. Scene objects and background (clutter) are represented by attributed relational graphs in which nodes represent objects of interest and arcs represent inter-object relations. Dynamic recognition/identification of nodes is accomplished by a belief/disbelief measure. Our experimental results with infrared images show improvements in natural scene object recognition over traditional image processing methods.			
20. DISTRIBUTION AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED <b>E</b>	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Robert N. Buchal		22b. TELEPHONE NUMBER (Include Area Code) (202) 767- 4939	
		22c. OFFICE SYMBOL NM	

DTIC FILE COPY

DTIC  
ELECTE  
NOV 26 1984

## Rule-based evidence accrual system for image understanding

Nelson Marquina

Artificial Intelligence/Image Understanding Section  
Honeywell Systems and Research Center  
2680 Ridgway Parkway, Minneapolis, Minnesota 55413

Abstract

The main function of an evidence accrual system for image understanding is to sequentially update information on scene objects based on new sensor data or on non-sensory information such as intelligence. This paper presents a concept for sequentially updating information on scene objects. Scene objects and background (clutter) are represented by attributed relational graphs in which nodes represent objects of interest and arcs represent inter-object relations. Dynamic recognition/identification of nodes is accomplished by a belief/disbelief measure. Our experimental results with infrared images show improvements in natural scene object recognition over traditional image processing methods.

Introduction

Honeywell Systems and Research Center, in a research contract with Air Force Office of Scientific Research (AFOSR), is currently working with real time air-to-ground sequences of infrared (IR) and range images with the objective of recognizing and identifying scene objects of interest.

Our approach to this problem is based on the notion of incremental acquisition of the scene model. Automatic object screener systems operate on video image frames, extract objects in a frame, and optimally classify these objects into objects and background based on their statistical and semantic features. The performance of the system (probabilities of false alarm and detection) depends on the quality of data and the image segmentors and the classifiers used by the system. The full potential of the segmentors and the classifiers is often not achieved because of severe system noise.

Misclassification may be reduced by examining the extracted objects and the classifier decisions on these objects over a sequence of image frames.<sup>1, 2</sup> This approach is useful and effective when noise in the image understanding system results in random noise in the processed image or random error in the feature values of the extracted objects, and the noise or the error is uncorrelated from frame to frame. When the image is noisy, an object may fail to meet the segmentation criteria of the system, resulting in a misclassification. When the feature values of the extracted objects are erroneous, there may be missed objects as well as false alarms. By accumulating information from one frame to the next regarding the locations and the feature values of the extracted objects, improved misclassification and detection can be achieved. Some reasons for the limitations of single-frame analysis approach are:

1. Objects in the scene may be occluded in any particular view.
2. Because of the high noise content of an air-to-ground IR image, it would be difficult to interpret all the scene segments.
3. Errors in analyzing and interpreting an image may create errors and inconsistencies in the scene description.

Our method involves using multiple views of the scene in a sequential manner. The different views are obtained via sensor motion (e.g., flying airplane or helicopter) and/or scene object motion (e.g., moving vehicles). A partial scene description using Attributed Relational Graphs (ARG's)<sup>3, 4</sup> is derived from each frame. As each successive frame is analyzed, the model of the scene is incrementally updated with information derived from the current frame. The model is initially a crude representation of the scene in which some objects may have been recognized, but most of them remain buried in a segment such as the case when background and object do not have high enough contrast, noticeable texture differences or objects have not moved significantly. An important aspect of dynamically updating the scene model as each frame is analyzed is the effective use of scene/image knowledge in the interpretation process,

particularly on methods and techniques for aggregating and mapping preliminary region, boundary, and/or surface information into higher-level descriptions.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Approved for public release;  
distribution unlimited.

84 11 19 033

The next section presents attributed relational graphs in the context of natural scene knowledge representation. Later on we discuss how to update the scene model by using belief/disbelief measures.

#### Scene Representation Using Graphs

Relational graphs have been used in several applications such as chemical structure description, picture encoding, relational data-base systems for pictures, network representation, switching theory, etc. The graph representation of images offers several powerful capabilities that are useful for image understanding such as the proper handling of the actual dimensionality and hierarchy of the images, the topological invariance, and the ability of having attributes (or features) attached to their nodes and arcs or branches.<sup>5,6</sup> Generally speaking, an Attributed Relational Graph (ARG) is a data structure defining an expected collection of objects, such as an outdoor scene, the expected visual attributes associated with the objects in the scene (each of which can have an associated ARG such as a syntactic decomposition), and the expected relations among them. For example, an outdoor scene can consist of classes "sky," "hill," "road," "vehicle," "tree," and "background." The class "vehicle" can have an ARG to represent the different types of vehicles expected in the scenario under observation. Furthermore, each type of vehicle is decomposed into its major parts such as "engine," "body," etc. The scene model represented by ARG's is sequentially updated by analyzing new frames. The interpretation of scene objects is dynamically accrued over time and convergence of interpretation yields the recognition of objects. The interpretation stage is performed by a rule-based system composed of production rules representing domain-specific knowledge about the scenario under observation. The information available to the rule-based system is composed of four classes: knowledge of form (shape, relative size, etc.), spectral characteristics (IR signatures, texture measures, etc.), plausible relations with other objects (convoy formation, on-road/off-road vehicles, etc.), and temporal profile (velocity, maneuverability, etc.). Interpretation rules relate image events to knowledge events by providing evidence for or against an object-hypothesis.

#### Evidence Accrual with Belief Measures

Decision smoothing techniques are utilized in object tracking systems to increase object tracking recognition confidence and decrease false alarm rate. Typically, statistical methods are employed in decision smoothing. These include accepting the mean (average) or mode (majority) target type over time, or using probabilistic models such as the Bayesian normal<sup>7</sup> or binomial,<sup>8</sup> to evaluate the likelihood of the various object types based on sequential observations. Simple statistical approaches, such as the mean and mode, suffer because they fail to account for auxiliary evidence of recognition accuracy which may be available as a by-product of the detection, segmentation, classification, and tracking processes. Schemes which attempt to use hypothesis testing theory [7], [8] to accumulate recognition confidence are unsatisfactory because they rely heavily upon assumptions of statistical models of the "decision population" which tend not to reflect the true nature of the observations. In particular, "evidence" for belief in various object classifications may come from several (statistically unrelated) sources. Thus, recognition of an object as a tank with 80% confidence does not generally imply that the confidence of other object types is less or equal to 20%. If classification was being done statistically on sets of features, it might well be the case that we attribute different confidences to the distributions of different object types in different feature spaces. In that case, classification results such as: "tank, 80%, APC, 40%," could make sense. Put another way, observation of "tank, 80%" is not necessarily equivalent to the conclusion "not tank, 20%." To utilize Bayes theorem to calculate the probability of "classification tank" given observed evidence (such as a set of features or a geometric arrangement of object components), requires that we know the probability of observing the evidence, given that the object is a tank. Yet the latter is precisely equivalent to the problem of modeling the object signature.

Lack of well-specified mathematical models confounds statistical accrual of classification confidence in many other areas of human endeavor including prediction of economic trends, weather forecasting, and medical diagnosis. Yet experts in these areas are often able to draw accurate conclusions on the basis of incremental (i.e., sequentially obtained) observations of evidence relevant to their conclusions. Shortliffe and Buchanan<sup>9</sup> have devised a method for incremental accrual of classification confidence which is motivated by the techniques employed by human experts in medical diagnosis. It was implemented originally in the MYCIN expert knowledge system<sup>10</sup>, and is now used routinely in the knowledge engineering field. The theory assumes that one can formulate approximations for a priori and conditional probabilities, but instead of treating these as strict statistical entities it uses them to determine

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)  
NOTICE OF SPONSORING AGENCY  
This technical report is the property of the  
approved for public release  
Distribution unlimited.  
MATTHEW J. KEEPER  
Chief, Technical Information Division

measures of "belief" and "disbelief." These belief measures are in turn used to define measures of confidence, and rules for incrementally updating both the belief and confidence measures are detailed.

The belief measures, as they are implemented in expert knowledge systems, are not time adaptive. That is, once some evidence for belief (or disbelief) in a hypothesis is accrued, its significance and numerical value never decreases. In the expert knowledge applications, such as medical diagnosis or mineral prospecting, this makes sense. During the time span of investigation, conditions are static; i.e., the disease or mineral deposits do not change their characteristics. However, this is not true of object signatures even over short time spans. Noise or low contrast may cause isolated frame mis-classification. It is desirable for the significance of this single observation to decrease over time. We have adapted the basic knowledge engineering belief measures to incorporate temporal context. Thus, hypotheses are formulated frame by frame, and time constants have been added to the incremental updating rules for absolute belief measures. Further temporal accrual of belief depends upon the gradient of the disbelief measure in the time domain, and vice versa.

#### Belief and Confidence Measures

Suppose we have a set of possible object types  $T_1, T_2, \dots, T_n$ , and a time sequence of frames through which we have tracked an object. Assume that in frame  $i$ , evidence  $e_{ij}$  is observed which supports the hypothesis,  $h_{ij}$ , that the tracked object is actually object type  $T_j$ . Assume also that confidence measure  $P(h_{ij})$  and  $P(h_{ij}/e_{ij})$  with

$$\begin{aligned} 0 &\leq P(h_{ij}) \leq 1 \\ 0 &\leq P(h_{ij}/e_{ij}) \leq 1 \end{aligned}$$

are calculated.  $P(h_{ij})$  is interpreted as the a priori confidence that in the  $i$ -th frame, the tracked object is  $T_j$ , and  $P(h_{ij}/e_{ij})$  is the conditional confidence that, after observing evidence  $e_{ij}$  in frame  $i$ , the tracked object is type  $T_j$ .

Define conditional measures of belief and disbelief that the tracked object in the  $i$ -th frame is type  $T_j$  by:

$$\begin{aligned} MB(h_{ij}, e_{ij}) &= \begin{cases} 1 & \text{if } P(h_{ij}) = 1 \\ \frac{\max[P(h_{ij}/e_{ij}), P(h_{ij})] - P(h_{ij})}{1 - P(h_{ij})} & \text{if } P(h_{ij}) \neq 1 \end{cases} \\ MD(h_{ij}, e_{ij}) &= \begin{cases} 1 & \text{if } P(h_{ij}) = 0 \\ \frac{\min[P(h_{ij}/e_{ij}), P(h_{ij})] - P(h_{ij})}{1 - P(h_{ij})} & \text{if } P(h_{ij}) \neq 0 \end{cases} \end{aligned}$$

Note that if  $MB(h_{ij}, e_{ij}) > 0$ , then,  $MD(h_{ij}, e_{ij}) = 0$  and vice versa. Thus, if the evidence,  $e_{ij}$ , increases belief in  $h_{ij}$ , then it cannot contribute to the disbelief in  $h_{ij}$ . Define absolute measures of belief and disbelief in the  $i$ -th frame by

$$\begin{aligned} MB(h_{ij}) &= \begin{cases} 0 & \text{if } MD(h_{ij}) = 1 \\ AMB(h_{i-1j}) + MB(h_{ij}, e_{ij})[1 - AMB(h_{i-1j})] & \text{if } MD(h_{ij}) \neq 1 \end{cases} \\ MD(h_{ij}) &= \begin{cases} 0 & \text{if } MB(h_{ij}) = 1 \\ BMD(h_{i-1j}) + MD(h_{ij}, e_{ij})[1 - BMD(h_{i-1j})] & \text{if } MB(h_{ij}) \neq 1 \end{cases} \end{aligned}$$

where  $A = 1 - MD(h_{i-1j}, e_{i-1j})$  and  $B = 1 - MB(h_{i-1j}, e_{i-1j})$  and

where the time constants,  $\alpha$  and  $\beta$  are fixed real numbers,  $0 < \alpha \leq 1$ ,  $0 < \beta \leq 1$ . Finally define the confidence that the tracked object is type  $T_j$  based on the evidence accumulated through the  $i$ -th frame by

$$CF(h_{ij}) = MB(h_{ij}) - MD(h_{ij})$$

$$\text{Then } -1 \leq CF(h_{ij}) \leq 1.$$

This follows because  $P(h_{ij})$  and  $P(h_{ij}/e_{ij})$  are between 0 and 1. So,  $0 \leq MB(h_{ij}, e_{ij}) \leq 1$  and  $0 \leq MD(h_{ij}/e_{ij}) \leq 1$ . From this we calculate  $0 \leq MB(h_{ij}) \leq 1$  and  $0 \leq MD(h_{ij}) \leq 1$ . In order to implement the sequential frame confidence scheme outlined before, it is necessary to define a priori single frame confidence  $P(h_{ij})$ , and conditional single frame confidence  $P(h_{ij}/e_{ij})$ , which depends on the evidence for  $h_{ij}$  (the hypothesis that in the  $i$ -th frame the tracked object is of type  $j$ ) observed in the  $i$ -th frame. These measures are necessarily dependent on the method used for object segmentation, since evidence for classification is derived during this process. We have developed ways of computing  $P(h_{ij})$  and  $P(h_{ij}/e_{ij})$  for syntactic and statistical classifiers.

Syntactic classification is based upon matching extracted target components to target models. Importance of components in the target model is given by weights  $w_k$ ,  $0 \leq w_k \leq 1$ , with  $\sum w_k = 1$ . Here,  $w_k$  is the significance of matching the  $k$ -th component in the target model. Note that "matching" must reflect the relative geometry of the target components. Then

$$\bar{c}_j = \sum_{\text{matched components}} w_k - \sum_{\text{missed components}} w_k$$

defines a preliminary confidence in the partial match of a candidate object to a target model. If only  $m$  of  $N$  components in the candidate object were matched in the model, then it makes sense to reduce confidence in the match proportionately. The confidence  $c_j$  are obtained by updating the initial value  $\bar{c}_j$  as a function of frame number and then obtain  $P(h_{ij}/e_{ij})$  by

$$P(h_{ij}/e_{ij}) = r(c_j)$$

where  $r(x)$  is a one-to-one non-linear rescaling function which maps  $[-1,1]$  onto  $[0,1]$  and also maps  $[1/2,1]$  onto  $[1/2,1]$ . Then  $e_{ij}$ , the evidence in the  $i$ -th frame that supports deciding in favor of target type  $T_j$ , is given by the match of the extracted components to those in the target model. Since we have a sequence of frames through which an object is tracked, we define the a priori confidence in the first frame to be  $P(h_{1j}) = 1/(n+1)$  where  $n$  object types are possible. The  $n+1$  represents the "probability" of "non-target," so that the tracked object necessarily matches one of the types. For the second frame, define  $P(h_{2j}) = P(h_{1j}/e_{1j})$ . In the  $i$ -th frame,  $i > 2$  define

$$P(h_{ij}) = \sum_{k=1}^{i-1} \alpha_{i-k-1} P(h_{kj}/e_{kj}).$$

The  $P(h_{ij})$  represents the weighted sum of historic evidence, up to the  $i$ -th frame, in favor of the target being classified as type  $T_j$ .

For the case of a statistical classifier, segmentation results in a set of feature values  $\mathbf{x} = (x_1, \dots, x_m)$  rather than extracted components. A parametric and a non-parametric scheme for calculating  $P(h_{ij}/e_{ij})$  have been developed. They are currently under experimental evaluation.

### Summary

An application of belief measures for decision smoothing over multiple frames has been presented. The same scheme generalizes to update confidence in any entity tracked through frames, providing "a priori" and "conditional" confidences can be attached to the entity at each frame. For instance, appearance of individual scene objects for syntactic classification is amenable to this process. Each object is matched against target models, and single frame confidence derived from the "goodness-of-match." Another example is confidence in tracked object velocity calculated between frame pairs. Single frame (pair) confidence is calculated from agreement (or lack of it) between current frame location and predicted location from the previous frame velocity vector.

A different set of applications of belief measures to image understanding exist which are not time dependent. (Setting the time factors for rule update, and  $\alpha$ , equal to zero eliminates time dependence in these rules.) Suppose that we wish to make a binary decision about a property of an object in an image, such as "round or not round", "match or not match a model or template", "component is merged or not", "component is a fragment or not", etc. Suppose that we have a set of presumably independent measures, each of which captures some aspect of the decision, and to each of which is associated a confidence measure between 0 and 1. For instance, a measure of roundness is given by smallest MSE from a circle, by the variance of the set of curvatures calculated at each point on the perimeter of the object, and also by the ratio of perimeter to pixel area. For each of these measures, a normalized scale of distance of the measure from that produced by a circle can be calculated, yielding a confidence measure. The problem is to accrue the strength of these various confidences to decide the total confidence in the decision "circle--not circle." A solution is to order the measures arbitrarily and treat them as sequentially obtained information (even though they can be obtained in parallel). Then the scheme outline in this paper (with time constants  $\alpha$  and  $\beta$  set to zero) can be applied to yield a single confidence measure. This measure can then be thresholded to determine the binary decision.

It is also possible to consider more complex combining schemes if more information, such as degree of dependence of measures, or independent confirmation of a measure, is available.

### Acknowledgment

This research was supported by the Air Force Office of Scientific Research under contract number F49620-83-C-0134.

### References

1. Panda, D.P., "Target screener system noise removal by interframe analysis," Proceedings: Image Understanding Workshop, Cambridge, MA., May 1978.
2. Herman, M., Kanade, T., and Kuroe, S., "Incremental acquisition of a three-dimensional scene model from images," IEEE Trans. PAMI, May 1984.
3. Shi, Q. and Fu, K., "Parsing and translation of (attributed) expansive graph languages for scene analysis," IEEE Trans. PAMI, Vol. PAMI-5, No. 5, September 1983.
4. Fu, Y.S., Syntactic Methods in Pattern Recognition. New York: Academic Press, 1974.
5. Eshera, M. and Fu, K., "A graph distance measure for image analysis," IEEE Trans. SMC, Vol. SMC-14, No. 3, June 1984.
6. Weymouth, T., Griffith, J., Hanson, A., and Riseman, E. "Rule-based strategies for image interpretation," Proc. Image Understanding Workshop, DARPA, June 1983.
7. Feldman, J.A. and Yakimovsky, Y., "Decision theory and artificial intelligence: I. A semantics based region analyzer," Artificial Intelligence No. 5, 1974, pp. 349-371.
8. Helland, A.R., Willet, T.J., Tisdale, G.E., "Applications of image understanding to automatic tactical target acquisition," Proc. of SPIE, Vol. 281, pp. 26-31, J.J. Pearson (Ed.), April 21-23, 1981.

9. Shortliffe, E.H. and Buchanan, B.G., "A model of inexact reasoning in medicine," *Mathematical Biosciences* 23, pp. 351-379, 1975.
10. Shortliffe, E.H., "MYCIN: A rule-based computer program for advising physicians regarding anti-microbial therapy selection," Doctoral Dissertation, Stanford University, October, 1974; Memo AIM-251, Stanford Artificial Intelligence Laboratory.

**END**

**FILMED**

**12-84**

**DTIC**